Niveau(x) d'enseignement : 0
Centre d'intérêt : 640 Numéro EDUQ : 9370293

Robert, Serge

Abduction, correction et apprentissage / par Serge Robert. -- Montréal :
Université du Québec à Montréal, Centre interdisciplinaire de recherche
sur l'apprentissage et le développement en éducation, 1992.
12 f. ; 28 cm. -- (Séminaire sur la représentation ; 69)

Notes bibliogr. au bas des p.

ISBN 2-921284-90-1

Vedette(s)-secondaire(s) : Université du Québec à Montréal. Centre
interdisciplinaire de recherche sur l'apprentissage et le développement
en éducation

Descripteur(s) : Ordinateur * Intelligence artificielle * Connaissance *
Théorie de l'apprentissage * Epistémologie * Recherche

Identificateur(s) de sujet : Signification * Sémantique

Le présent texte veut approfonuir le débat autour de l'ordinateur
qu'on doit considérer ou non comme une machine pensante et cognitive,
comme une "intelligence artificielle". Il présente les points de vue des
chercheurs sur la possibilité de rendre les ordinateurs des machines
autant sémantiques que syntaxiques. Il discute de la signification et de
la représentation de la signification. Il incite à poursuivre la
recherche en intelligence artificielle, à représenter l'activité
sémantique et les autres fonctions cognitives humaines et à envisager la
possibilité de les intégrer à l'ordinateur. Il évoque les différentes
conceptions qui ont cours sur la nature des machines, conceptions
dépendantes des relations entre la matière et l'esprit. Il propose une
théorie interactionniste de la connaissance qui peut suggérer des pistes
au développement de l'intelligence artificielle et des sciences
cognitives. Et il suggère d'orienter la recherche en intelligence
artificielle du côté des opérations d'abduction et de correction qui sont
liées à des propriétés spécifiques à l'être humain. Ces opérations font
voir les limites de l'ordinateur qui jusqu'à maintenant est considéré
comme une machine déductive et non comme une machine cognitive
rationnelle.

Type de document :

CENTRE INTERDISCIPLINAIRE DE RECHERCHE SUR
L'APPRENTISSAGE ET LE DÉVELOPPEMENT EN ÉDUCATION

cirade

Université du Québec à Montréal

9370293

Abduction, correction et apprentissage

par Serge Robert
Département de philosophie
Université du Québec à Montréal

Séminaire sur la Représentation

14 février 1992 - n° 69

CIRADE, UQAM, novembre 1992

À travers l'histoire, de grandes inventions techniques ont contribué à diminuer l'effort humain tout en accroissant l'efficacité du travail. Pour arriver à ce résultat, on a dû produire, souvent avec une ingéniosité remarquable, des outils qui se sont avérés être des substituts plus efficaces à des activités humaines: ainsi, par exemple, la bicyclette et l'automobile ont supplanté la vitesse de la marche, comme divers types de moteurs ont accru la puissance de l'effort humain. La plupart de ces réalisations techniques ont été des substituts d'activités physiques (des technologies du corps), mais certaines, comme l'invention de la machine à calcul, ont remplacé des activités mentales. Au cours des dernières décennies, les ordinateurs se sont présentés comme de prodigieuses technologies de l'esprit, c'est-à-dire comme des substituts très efficaces d'activités mentales. Ceci est d'autant plus vrai avec la réalisation, dans le domaine de l'intelligence artificielle, des programmes informatiques experts, qui rendent l'ordinateur capable de faire des diagnostics et de rendre des décisions qui jusqu'à maintenant n'étaient réservées qu'à des experts humains. Ainsi, des programmes médicaux experts rendent des diagnostics à partir d'une énumération de symptômes du patient, comme des programmes militaires orientent la bombe en réaction à des données géographiques et météorologiques. Relativement à l'expert humain, le programme expert a même l'avantage de rendre des jugements plus rapides que lui, et également plus sûrs, parce qu'exempts des sources humaines d'erreur, comme la distraction, l'oubli, la fatigue ou le doute. Peut-on alors véritablement parler d'une «intelligence artificielle»? L'ordinateur peut-il être considéré comme une machine pensante et cognitive? Ce type de questions a récemment donné lieu à un débat entre John Searle et Paul Churchland, dans la revue *Scientfic American*, débat traduit en français dans la livraison de mars 1990 de *Pour la Science*[1]. Le présent texte[2] vise à reprendre ce débat, de façon à tenter de l'approfondir davantage et à soumettre à ce sujet des hypothèses nouvelles.

1. Y a-t-il un sinophone dans la chambre chinoise?

Searle expose sa position dans son célèbre argument de la chambre chinoise[3]. L'argument se présente comme suit: imaginons une pièce fermée, sans lien avec l'extérieur, où un individu se fait présenter sur une table des successions de cartes portant chacune un signe, lequel est en fait, sans qu'il ne le sache, un idéogramme chinois. Assis devant cette table, l'individu n'a à sa disposition qu'un panier de cartes idéogrammatiques et on lui a appris à réagir à chaque succession déterminée de cartes qu'on lui présente par une succession précise de cartes qu'il doit sortir de son panier et qu'il doit étaler sur la table, sous les idéogrammes qu'on lui a présentés. L'apprentissage qu'on lui a imposé est tel que si on invite un observateur qui

---

1 Voir Searle, J., «L'esprit est-il un programme d'ordinateur?» et Churchland, P., «Les machines peuvent-elles penser?», *in Pour la Science*, mars 1990, pp. 38-44 et pp. 46-53.

2 Le présent texte explicite certaines idées que l'on retrouve dans la section 2 du chapitre VII de Robert, S., *Découverte, justification et cognition: une épistémologie rationaliste interactionniste*, manuscrit en voie de publication.

3 Voir Searle, J., «*Minds, Brains and Science*», Cambridge, Mass., Harvard University Press, 1984, pp. 32-35 et «L'esprit est-il un programme d'ordinateur?», 1990.

connaît le chinois à entrer dans la salle et à observer les réactions de l'individu aux successions d'idéogrammes qu'on lui présente, cet observateur va juger, s'il ne connaît pas les conditions de l'apprentissage imposées à l'individu, que l'individu répond en chinois avec pertinence à des questions qui lui sont posées en chinois, et donc, qu'il a appris et connaît le chinois. Par contre, celui qui connaît les conditions de cet apprentissage, sait que l'individu n'a appris qu'à réagir à des successions de signes par d'autres successions de signes, sans qu'il ne sache que ce qu'il fait peut être interprété comme des réponses en chinois à des questions posées en chinois. Pour Searle, l'individu de la chambre chinoise ne connaît pas le chinois et ne le parle pas, parce qu'il ne connaît que la syntaxe du chinois, sans en connaître la sémantique. L'ordinateur serait selon Searle dans la même situation que cet individu, c'est-à-dire qu'il serait une machine qui syntaxise sans sémantiser, qui ne parlerait donc pas un langage et qui, bien qu'utile, ne pourrait pas être considérée comme intelligente[4]. Pour Searle, tout ordinateur ne peut être que syntaxique sans être sémantique: pour pouvoir sémantiser, il faudrait qu'il soit capable de désirs et de croyances, lesquels sont considérés par Searle comme des conditions intentionnelles nécessaires à la production d'actes de signification. Or pour être porteur de désirs et de croyances, il faut, d'après Searle, posséder un cerveau, ou une substance équivalente, dont émerge un esprit, ce qui n'est pas le cas pour un ordinateur, si complexe soit-il[5]. Parler d'«intelligence artificielle» serait donc un abus de langage.

Paul Churchland s'est opposé à la position searlienne, en soutenant que nous ne pouvons a priori considérer l'intelligence artificielle comme impossible et que, bien que les ordinateurs aient été jusqu'à maintenant des machines plus syntaxiques que sémantiques, il est tout à fait possible de les rendre sémantiques[6]. Pour Churchland, les ordinateurs traditionnels ont été conçus dans une perspective représentationniste, où l'information est précisément organisée comme un système syntaxique. Pour leur permettre d'accomplir des actions plus complexes et plus semblables à l'activité cognitive humaine, il faut plutôt les construire dans une perspective connexionniste, où l'information est organisée selon des réseaux complexes, comparables aux connexions synaptiques qui relient les neurones dans le cerveau humain[7]. Ainsi, on peut tout à fait raisonnablement concevoir que l'individu de la chambre chinoise, qu'il soit un ordinateur ou un humain, peut se faire inculquer des fonctions d'interprétation des idéogrammes, où à chaque idéogramme est associée une référence, et devenir ainsi une machine tout autant sémantique que syntaxique. En plus de savoir associer des idéogrammes entre eux, cette machine saurait associer des idéogrammes à des événements, des objets, des propriétés, des classes, etc.

---

4 Searle, J., 1984, p. 39 et «L'esprit est-il un programme d'ordinateur?», 1990.
5 Searle, J., 1984, p. 41 et «L'esprit est-il un programme d'ordinateur?», 1990.
6 Churchland, P.M., «Les machines peuvent-elles penser?», *in* 1990.
7 *Ibid.*

L'enjeu de la controverse Searle-Churchland repose sur la conception que l'on a de la sémantique et, plus particulièrement, sur la possibilité que l'on a de la représenter syntaxiquement. Le point de vue de Churchland suppose que la dimension sémantique du langage est réductible à des fonctions du langage-machine, c'est-à-dire à un type de représentation syntaxique particulier, ou à tout le moins à un système de connexions spécifiques, où des fonctions associent des signes à des éléments de référence, plutôt que d'associer des signes à d'autres signes, comme le font habituellement des règles syntaxiques. Au contraire, la conception searlienne considère qu'une telle réduction syntaxique est insuffisante pour reproduire la richesse de l'activité sémantique et que l'acte sémantique est impossible sans que des désirs et des croyances, jusqu'à maintenant réservés à des êtres vivants, ne rendent possible cet acte.

2. Signification et représentation de la signification

On peut profiter du fait que Tarski a produit une sémantique formelle et une théorie formelle de la vérité pour les langues formalisées[8], pour en appliquer les résultats au présent débat. On peut alors dire dans un premier temps que la réussite tarskienne donne en partie raison à Churchland, puisqu'elle représente syntaxiquement l'interprétation sémantique d'une expression comme étant une fonction entre cette expression et une entité externe au langage auquel l'expression appartient[9]. Là où cependant la thèse de Searle apparaît juste, c'est quand, pour exprimer l'entité de référence de cette fonction d'interprétation, on doit la représenter comme une expression d'un niveau de langage différent du niveau où se trouve l'expression dont elle est la référence. Pour ce faire, on distingue le niveau de langage de l'expression dont on donne la référence, en en faisant mention entre guillemets, du niveau de langage de la référence de cette expression, dont on fait usage sans guillemets. Tel est par exemple le cas dans le célèbre exemple tarskien: l'expression «La neige est blanche» est vraie, si et seulement si la neige est blanche (dans les faits)[10]. Référer, c'est donc associer une expression d'un langage à une entité qui n'appartient pas à ce langage, tandis que représenter l'acte de référence c'est associer la mention de l'expression à l'expression elle-même. Ceci montre bien que représenter la signification dans une règle syntaxique constitue une réduction linguistique d'un acte qui consiste, au contraire, à associer les expressions du langage à un extérieur qui ne lui appartient pas. Autrement dit, comme le montre bien Tarski, aucun langage ne possède sa propre notion de vérité et n'est en mesure de nous dire ce qui est vrai et ce qui est faux en son sein, sa notion de vérité lui étant métalinguistique. Faire la sémantique d'un langage-objet ne peut ainsi se faire qu'au sein d'un métalangage extérieur à ce langage-objet et apte à faire mention des expressions du langage-objet. Ainsi, le désir et la croyance qui, pour Searle, associent dans l'acte

8 Voir entre autres Tarski, A., «Le Concept de la vérité dans les langues formalisées» *in Logique, sémantique et métamathématiques*, G.G. Granger (éd.), Paris, Armand Colin, 1972.

9 *Ibid.*

10 *Ibid.*

sémantique le langage à l'expérience extra-linguistique sont, en dernière instance, irréductibles à une représentation dans un langage.

La notion de sémantique apparaît ici dans deux sens différents: dans un sens faible, la sémantique est faite de fonctions d'interprétation, qui associent entre elles des expressions de niveaux de langage différents (certaines dont on fait mention et d'autres dont on fait usage); dans un sens fort, la sémantique est une activité complexe qui n'est que représentée et non réellement présente dans ces fonctions d'interprétation, puisqu'elle relie des expressions à des entités extra-linguistiques, et que de cette façon elle n'est pas réductible à un langage ou au fonctionnement d'une machine. Au sens fort, la sémantique est une activité qui associe des usages d'expressions et des entités extra-linguistiques; au sens faible, elle est une représentation de cette activité qui associe des mentions d'expressions à des usages d'expressions. Mon détour par Tarski montre que c'est donc la sémantique au sens faible d'une représentation qui m'apparaît pouvoir être présente dans une machine comme les ordinateurs que nous possédons maintenant, et non la sémantique au sens fort de l'activité qu'on représente. Sur le plan ontologique, la position de Searle est donc la plus appropriée, puisqu'elle m'apparaît reposer sur une différence fondamentale entre l'esprit humain, qui réussit à relier des expressions à des événements physiques ou psychologiques, et la machine, telle qu'on la connaît, qui ne réussit qu'à relier des expressions entre elles.

3. Le postulat méthodologique de Churchland

Sur le plan méthodologique, la position de Churchland m'apparaît cependant être celle qu'il faut suivre en ce qui concerne le domaine de l'intelligence artificielle. Alors que le point de vue searlien peut porter le philosophe à bouder l'informaticien et la recherche en intelligence artificielle, leur objectif ultime apparaissant utopique, celui de Churchland incite au contraire à poursuivre cette recherche et à associer le travail du philosophe à celui de l'informaticien. En travaillant à représenter l'activité sémantique et les autres fonctions cognitives humaines, nous ne pouvons arriver qu'à deux types possibles de résultats, qui peuvent tous deux constituer un progrès scientifique. D'une part, on peut réussir à représenter dans une machine l'activité sémantique, obtenir ainsi une machine technologiquement utile et mieux comprendre en quoi l'activité sémantique réelle est différente de cette représentation. D'autre part, on peut, ce qui apparaît présentement difficilement pensable, compte tenu de mon argumentation pro-Searle, mais pas nécessairement impossible, complexifier la machine au point d'arriver à réaliser l'objectif de Churchland, à intégrer dans la machine une activité sémantique proprement dite, et alors mieux comprendre, à travers l'étude du fonctionnement de cette machine, l'activité cognitive humaine. D'une manière ou de l'autre, le développement de l'intelligence artificielle (qui vise à rendre les machines cognitives) peut donc contribuer au développement d'une science cognitive (qui cherche à expliquer la connaissance humaine comme étant semblable au fonctionnement d'une machine): si on réussit à produire une machine vraiment intelligente, on

peut mieux connaître l'intelligence humaine en étudiant cette machine; si, au contraire, on ne réussit pas, comme ce fut le cas jusqu'à maintenant, on peut approfondir notre connaissance de l'esprit humain en le contrastant relativement à la machine, comme Searle a pu le faire et comme je le fais maintenant. De toute façon, le progrès de la technologie informatique comme le progrès de la connaissance épistémologique du fonctionnement de la connaissance exigent, en ce qui concerne l'intelligence artificielle, de nous rallier à la position méthodologique de Churchland et de faire comme si une machine pouvait penser et connaître.

Je ne crois pas que le débat entre Searle et Churchland puisse être tranché catégoriquement, puisqu'il dépend de la conception que l'on a de ce qu'est et de ce que peut être une machine. Ainsi, considérer comme Churchland qu'on doive faire l'hypothèse qu'une machine pensante est réalisable n'est possible que dans la mesure où on parle de machines qui n'existent pas présentement et qui seraient suffisamment différentes de nos ordinateurs actuels pour produire une véritable activité cognitive, plutôt que de simplement en représenter certains aspects. Soutenir comme Searle qu'un ordinateur ne peut penser est juste, si on se réfère non seulement aux ordinateurs actuels, mais aussi à notre concept même de ce que peut être un ordinateur, cette fois au sens d'une machine qui peut au mieux représenter, et non effectuer, une activité cognitive. Cela dit, personne n'est en mesure de soutenir contre Churchland qu'il nous est impossible de produire une machine suffisamment différente de nos ordinateurs actuels et suffisamment complexe pour effectuer une activité cognitive, même si une telle machine est actuellement inconcevable et nous apparaît, par le fait même, très improbable. C'est en dernière instance un tel argument qui invalide toute tentative de jeter le discrédit sur l'hypothèse méthodologique de Churchland. Searle reconnaît cette possibilité d'une machine très différente de nos ordinateurs et aux propriétés semblables à celles du cerveau et, ce faisant, sa position n'est pas entièrement incompatible avec celle de Churchland. La réalisation d'une telle machine rejoindrait certaines propriétés de la reproduction humaine: quand nous faisons un enfant, nous créons, entre autres, un système cognitif, mais sans savoir comment nous y arrivons, tandis que lorsque nous créons un ordinateur doté de programmes experts, nous savons ce que nous faisons sans arriver à en faire une machine cognitive. La réalisation d'une machine vraiment cognitive serait donc un dépassement de l'opposition de l'esprit et de la machine.

4. L'esprit et la machine

Les différentes conceptions sur la nature des machines sont elles-mêmes dépendantes de conceptions sur les relations entre la matière et l'esprit. Ainsi, un dualisme radical entre matière et esprit, que j'appelle un mentalisme fort, soutient que l'esprit a une existence complètement autonome relativement à la matière et qu'il est totalement irréductible à la matière: une telle conception ne s'intéresse même pas au débat dont il est question ici et considère d'entrée de jeu qu'une machine, comme tout autre objet matériel, ne peut réussir à avoir des comportements réservés à l'esprit, comme la production de significations et de connaissances. Conformément à

mon argumentation de la section précédente, que le dualisme soit vrai ou faux, il est de toute manière épistémologiquement stérile.

Contre le dualisme, le matérialisme réductiviste soutient que l'esprit est réductible à de la matière, en l'occurrence au cerveau, qu'un comportement mental est un comportement matériel complexe et donc qu'il est possible qu'une machine complexe, dont le cerveau serait un exemple, effectue des activités mentales. Sans aller jusqu'à épouser ce réductivisme d'un point de vue ontologique, Churchland le met de l'avant d'un point de vue méthodologique, dans ce qu'il appelle un matérialisme méthodologique[11]. L'opposition entre le dualisme et le réductivisme porte essentiellement sur la frontière entre matière et esprit et sur le degré de puissance créatrice que l'on attribue à la matière: ainsi le matérialisme réductiviste donne à la matière une capacité comportementale que le dualisme ne lui reconnaît pas. L'enjeu pour un réductiviste méthodologique comme Churchland est alors de déterminer quelle complexification matérielle est nécessaire et suffisante à ajouter à la machine pour qu'elle devienne apte à penser et à connaître?

La position de Quine sur les relations matière/esprit est, quant à elle, une version épistémologique du matérialisme réductiviste, à savoir le physicalisme, lequel consiste à n'accepter comme scientifiques que les explications en termes de comportements physiques, de sorte que les événements psychologiques ne soient acceptables qu'à la condition d'être descriptibles et explicables en termes physiques[12]. Ce physicalisme amène Quine à réduire la signification à des stimuli (savoir le sens d'une expression se réduit à savoir dans quels contextes il est approprié de l'utiliser) et à traiter la traduction entre deux locuteurs de langues différentes comme étant fondamentalement indéterminée (nous n'avons pas de critère objectif pour considérer une expression d'une langue comme équivalente à une expression d'une autre langue)[13]. À partir de là, on peut déduire que la position quinienne sur l'argument de la chambre chinoise serait que l'indétermination de la traduction et le traitement de la signification comme des stimuli sont tels qu'un locuteur humain parlant le chinois doit être considéré comme ne faisant rien de plus que l'individu de la chambre chinoise, comme ne sachant qu'associer des idéogrammes-réponses à des idéogrammes-questions. Dans cette perspective, non seulement l'ordinateur peut penser, mais il peut le faire avec les moyens informatiques actuels, simplement par des règles syntaxiques appropriées, sans qu'il soit nécessaire de chercher à lui faire exécuter des comportements sémantiques d'un statut différent et plus complexe que celui des opérations syntaxiques.

Dans ce débat, une troisième position intermédiaire entre le dualisme et le réductivisme me paraît fertile, à savoir un matérialisme émergentiste et un mentalisme faible, selon lesquels

---

11 Churchland, P.M., «*Matter and Consciousness*», Cambridge, Mass., MIT Press, 1984, pp. 96-98.

12 Quine, W.V.O., entre autres, «*Theories and Things*», Cambridge, Mass., Harvard University Press, 1981, p. 98.

13 Quine, W.V.O., entre autres «*Pursuit of Truth*», Cambridge, Mass., Harvard University Press, 1990, pp. 47-49.

l'esprit a ses caractéristiques propres, irréductibles au comportement de la matière, mais n'existe que comme complexification qui émerge de la matière. Pour le matérialisme émergentiste, la matière complexe du cerveau, ou de quelque autre matériau semblable dont nous ignorons s'il peut exister, est un support nécessaire de l'esprit. Cette position rejoint le point de vue de Searle, pour qui l'esprit est irréductible au cerveau, sans cependant exister en dehors d'un tel support, dont la complexité matérielle rend possible l'émergence de l'esprit[14].

Mon mentalisme est faible non seulement parce qu'il ne postule pas l'auto-subsistance de l'esprit, mais aussi parce qu'il est méthodologique. La différence entre ma position et celle de Searle est que pour moi l'irréductibilité de l'esprit à des propriétés matérielles ne s'appuie pas sur une opposition de principe, mais sur l'état actuel de l'histoire des sciences, où les sciences physiques ne sont pas en mesure de rendre compte des comportements humains complexes que nous révèlent les sciences humaines et plus particulièrement les sciences cognitives. Nous avons des expériences physiques, comme quand nous nous heurtons à la matérialité des choses, et des expériences psychologiques, comme lorsque nous rêvons ou construisons des concepts, et la connaissance est essentiellement une interaction étroite de ces deux types d'expériences. Compte tenu de l'état historique de nos connaissances, ces deux types d'expériences nous apparaissent distincts, et l'expérience empirique, que nous vivons dans la perception, nous apparaît comme le lieu privilégié de rencontre de l'expérience physique et de l'expérience psychologique. Dans cet ordre d'idées, j'ai développé[15] une théorie rationaliste interactionniste de la connaissance qui consiste à soutenir que la connaissance est un comportement d'harmonisation progressive et interminable entre notre expérience psychologique et notre expérience physique, par corrections successives de nos représentations psychologiques à la lumière de ce que notre perception nous révèle de notre expérience physique. Si un jour la science physique nous permettait d'expliquer totalement notre expérience psychologique en termes physiques, ce qui nous paraît présentement inconcevable, parce que nous sommes bien loin d'avoir réussi à expliquer le passage de la matière à l'esprit, alors les sciences cognitives devraient être remplacées par une physique de la connaissance.

Pour l'instant, et probablement pour très longtemps encore, sinon pour toujours, la position qui m'apparaît la plus rationnelle est un matérialisme réductiviste méthodologique à la Churchland en ce qui concerne l'intelligence artificielle, de façon à produire des machines de plus en plus performantes et à faire progresser notre connaissance des fondements matériels de la connaissance et des relations de la matière et de l'esprit, et un matérialisme émergentiste méthodologique, proche de Searle, en matière de sciences cognitives, pour faire progresser notre connaissance de la spécificité de l'esprit et de son rôle dans l'action humaine et dans le fonctionnement de la connaissance. Produire des machines, c'est humaniser la nature, au sens

---

14 Searle, J., 1984, chapitre 2, surtout p. 39.

15 Voir Robert, S., *Découverte, justification et cognition: une épistémologie rationaliste interactionniste*, manuscrit de livre, en voie de publication.

d'un accroissement de l'emprise humaine sur la nature. La plupart des inventions techniques (les technologies du corps) ont accru notre emprise physique sur la nature, tandis qu'avec le développement de l'intelligence artificielle, c'est notre emprise mentale sur la nature qui prend de l'importance. En travaillant à connaître les mécanismes de fonctionnement de la connaissance humaine, les sciences cognitives essaient de représenter cette connaissance comme étant semblable au fonctionnement d'une machine. Entre l'humain et son environnement naturel, les sciences cognitives parcourent donc le chemin inverse de l'intelligence artificielle. En somme, le développement technologique des machines et le progrès théorique des connaissances apparaissent comme le résultat historique de l'interaction complexe entre l'humain et son environnement. On peut alors penser que la distance entre l'intelligence artificielle et les sciences cognitives va désormais progressivement diminuer: l'avenir de l'intelligence artificielle nous donnera des machines qui ressembleront de plus en plus au comportement cognitif humain et les sciences cognitives vont de plus en plus expliquer le fonctionnement de la connaissance humaine comme étant semblable au fonctionnement algorithmique d'une machine. Pour arriver à ces résultats, le matérialisme réductiviste à la Churchland s'avère le postulat méthodologique le plus utile à l'intelligence artificielle, tout comme le matérialisme émergentiste, proche de Searle, peut jouer le même rôle pour les sciences cognitives. La controverse Searle-Churchland m'amène donc à une position intermédiaire qui m'apparaît plus stimulante pour le progrès scientifique. Autrement dit, les sciences cognitives ne peuvent progresser que si elles postulent qu'il y a dans le fonctionnement mental quelque chose de plus complexe que ce que la nature nous donne à connaître, et l'intelligence artificielle ne peut progresser que si elle postule à son tour que ce plus complexe est quand même réductible au fonctionnement matériel d'une machine. Cette dialectique méthodologique permet de penser que la fusion de l'intelligence artificielle et des sciences cognitives, ou la réalisation d'une physique de la connaissance, ne sont pour la science que des idéaux régulateurs probablement inaccessibles.

5. La théorie interactionniste de la connaissance

Pour suggérer des pistes au développement de l'intelligence artificielle et des sciences cognitives, je peux rappeler ici les principaux résultats auxquels je suis arrivé dans ma théorie interactionniste de la connaissance humaine[16]. Selon cette théorie, la connaissance commence dans la perception, qui, en tant qu'expérience empirique, est une rencontre de notre expérience physique et de notre expérience psychologique. La perception est une construction sur un donné sensoriel: cette construction est plus précisément abductive, au sens où elle sélectionne des différences sensorielles de façon à nous faire percevoir des objets, elle induit à partir de ressemblances sensorielles pour nous faire percevoir des propriétés et des relations sur les objets et, à partir de là, elle déduit de façon anticipative des événements empiriques, comme étant des liens entre des objets et des propriétés ou relations perceptibles. Les éléments de notre environnement que nous retenons par nos opérations de sélection dépendent de nos intérêts

16 *Ibid*

pragmatiques de connaissance et le contenu de nos inductions est relatif aux régularités qui sont spécifiques à notre environnement. L'imagination intervient ensuite de façon créative sur nos résultats perceptifs en les réorganisant, de façon à produire des entités imaginaires (comme des objets, des propriétés, des relations, ou des événements imaginaires). La représentation verbale intervient ensuite pour représenter les résultats de la perception et de l'imagination dans le langage verbal. Le langage verbal est ainsi, à son tour, obtenu par abduction, c'est-à-dire par sélection, par induction, par déduction et, parfois, par réorganisation imaginaire. Alors que l'induction verbale permet de passer d'un niveau de langage inférieur à un niveau supérieur, la déduction nous fait passer, en sens inverse, d'un niveau supérieur à un niveau inférieur. Ainsi, la connaissance humaine se joue principalement sur trois niveaux de langage: un premier niveau dont la fonction est descriptive, un deuxième niveau, qui est explicatif et, enfin, un troisième niveau, lequel est justificatif (ou apte à expliquer des explications). Alors que le niveau explicatif est celui où, à travers des définitions et des hypothèses causales, on retrouve le discours scientifique, le niveau justificatif est celui des fondements métascientifiques de la connaissance, comme ce qu'on retrouve dans les lois logiques ou mathématiques, dans les hypothèses de la physique théorique ou dans les théories épistémologiques. L'évolution rationnelle de la connaissance consiste dans l'élimination de problèmes d'inconsistance dans le système total de la connaissance, certaines de ces inconsistances étant horizontales, ou internes à un niveau de langage, d'autres étant verticales, ou entre deux niveaux de langage successifs (par exemple, entre le descriptif et l'explicatif, ou entre l'explicatif et le justificatif). Éliminer une inconsistance constitue un progrès cognitif et les inconsistances verticales peuvent être éliminées par des opérations de correction, qui sont soit des exclusions extensionnelles (ou ajustements par le niveau discursif du bas) ou des remaniements intensionnels (ou ajustements par le niveau discursif du haut). En résumé, selon cette théorie rationaliste interactionniste, la connaissance est une activité de représentation psychologique de notre expérience physique, par la médiation de notre expérience empirique, par le biais d'opérations abductives (sélection, induction, réorganisation imaginaire et déduction) et d'opérations correctives (exclusion extensionnelle et remaniement intensionnel).

6. Les limites de l'ordinateur et les propriétés d'une machine cognitive

Si on compare les résultats auxquels ma théorie interactionniste de la connaissance nous a conduits avec le fonctionnement des ordinateurs, on doit conclure que, pour qu'un ordinateur soit considéré comme une machine cognitive, il doit être à la fois capable d'abductions et de corrections. Or nos ordinateurs sont surtout des machines déductives. Si on réussit à leur faire produire les autres opérations abductives, à savoir des sélections, des inductions et des recombinaisons d'informations, on ne peut parler de véritable abduction cognitive, parce que la sélection ne se fait pas en fonction d'intérêts à réaliser et l'induction ne se fait pas en fonction des régularités de l'environnement. Ainsi, sans désir d'intervention dans son environnement et sans sensibilité inductive relativement à cet environnement, l'ordinateur ne peut faire

d'abductions vraiment cognitives. Sans une interaction avec son environnement qui serait de même nature que celle que nous avons dans notre lutte pour survivre, l'ordinateur est incapable d'associer des informations à des entités extérieures à lui: il est donc précisément incapable, conformément à la thèse searlienne, d'activité sémantique et d'activité cognitive. Sans désir et sans vulnérabilité informationnelle relativement à son environnement, l'ordinateur ne signifie pas, il ne veut rien dire, et donc, il ne connaît pas. Mon ordinateur ne soutient pas les thèses que j'exprime à travers lui en rédigeant ce texte. Dans la mesure où l'abduction cognitive, qui permet au langage humain de produire de la signification et de la connaissance, commence au niveau de la perception, il n'est pas surprenant que la perception soit précisément une fonction très difficile à réaliser en intelligence artificielle, probablement celle où le progrès des réalisations technologiques est le plus lent: c'est dans la perception qu'est la forme la plus fondamentale de l'activité cognitive, là où notre expérience psychologique et notre expérience physique se rencontrent. C'est aussi dans la perception que l'intelligence rudimentaire des animaux commence à se manifester.

Si l'ordinateur était capable d'abductions cognitives, sa connaissance ne pourrait pas pour autant être considérée comme scientifique ou rationnelle, s'il ne réalisait pas de corrections. S'il était capable d'abductions cognitives, l'ordinateur prêterait le flanc à des inconsistances. Or les inconsistances rendent l'ordinateur dysfonctionnel: sa structure déductive lui fait tirer n'importe quelle conclusion à partir de prémisses inconsistantes. Pour lui permettre de connaître scientifiquement, de résoudre des problèmes cognitifs et de progresser, il faudrait le rendre capable de suspendre temporairement son activité déductive, de façon à lui permettre de réorganiser son information par des corrections. Un micro-ordinateur, par exemple, peut être considéré comme un système syntaxique à trois niveaux d'informations, comparables aux niveaux du discours cognitif: les documents sont de même niveau que le langage descriptif, le logiciel effectue des opérations sur ce niveau et est comparable au niveau du langage explicatif, enfin, le système fait des opérations sur le logiciel et a ainsi un rôle comparable au niveau du langage justificatif. Un ordinateur capable de connaissance rationnelle serait donc capable par lui-même de modifier son information, de façon à éliminer des inconsistances horizontales internes à un système, à un logiciel ou à un document, et d'éliminer des inconsistances verticales entre un système et un logiciel, ou entre un logiciel et un document. Tant qu'il ne procède pas par des processus abductifs corrigibles, l'ordinateur n'est pas une machine cognitive rationnelle, ce qui ne l'empêche pas toutefois d'être un instrument déductif fort utile.

Compte tenu de mes positions épistémologiques, c'est, en conclusion, du côté des opérations d'abduction et de correction, telles que présentées ici, que je suggère d'orienter la recherche en intelligence artificielle. À ces opérations d'abduction et de correction sont liées des propriétés spécifiques à l'esprit humain comme: un désir sélectif d'intervenir dans son environnement et une sensibilité inductive à cet environnement, une aptitude à la représentation verbale des événements, de manière à les décrire, à les expliquer et à justifier ces explications,

une aptitude formelle à formuler du vrai, du faux, du tautologique, du contradictoire et du paradoxal, une aptitude métadiscursive à mentionner, à nier, à mentir et à blaguer, une capacité de jeu, de liberté, de comportement éthique et de comportement esthétique, l'existence d'une activité psychologique intense et d'une conscience réflexive, une capacité de se donner des critères de vérité, de mathématiser, d'anticiper sa propre mort, de croire à des divinités, de sombrer dans la psychose... Devant une telle énumération, il apparaît rassurant de penser que la réalisation de machines vraiment cognitives n'est qu'un idéal régulateur inaccessible, étant donné qu'on peut difficilement anticiper tous les problèmes sociaux, politiques et éthiques que nous poserait l'existence de telles machines...

Des exemplaires du présent document sont en vente à l'adresse suivante:

Centre Interdisciplinaire de Recherche sur
l'Apprentissage et le Développement en Education
Université du Québec à Montréal
C.P. 8888, Succ. A,
Montréal, Québec
H3C 3P8

Publié par le Centre Interdisciplinaire de Recherche
sur l'Apprentissage et le Développement en Education
Université du Québec à Montréal
Conception graphique: Pierrette [illegible]